



清华大学高等研究院

Institute for Advanced Study, Tsinghua University

学术报告

- Title:** Why are recurrent neural networks NOT good at natural language processing?
- Speaker:** Huitao Shen 沈汇涛 (MIT)
- Time:** 3:00pm, Friday, June 21, 2019
- Venue:** Conference Hall 322, Science Building, Tsinghua University

Abstract

Sequence models assign probabilities to variable-length sequences such as natural language texts. The ability of sequence models to capture temporal dependence can be characterized by the temporal scaling of correlation or mutual information. In this talk, I will show the mutual information of recurrent neural networks (RNNs) decays exponentially in temporal distance, analytically in linear Gaussian RNNs and empirically in nonlinear RNNs such as long short-term memories. On the other hand, self-attentional models like Transformers can capture long-range mutual information more efficiently, making them preferable in modeling sequences with slow power-law mutual information, such as natural languages and stock prices. The connection of the results with statistical mechanics will be discussed.

Reference: arXiv: 1905.04271